

Serial No. 10/059,421
Docket No. YOR920010126US2

2

AMENDMENTS TO THE CLAIMS:

1. (Currently amended) A system for identifying genes, comprising:
a pattern database comprising patterns of amino acids;
an input device for inputting a DNA sequence; and
a processor which:
translates an open reading frame (ORF) of said DNA sequence into an amino acid translation; and
locates in said amino acid translation occurrences of said patterns a pattern from said pattern database in said DNA sequence to determine whether said open reading frame includes identify a putative gene in said DNA sequence.
2. (Currently amended) The system according to claim 1, wherein said processor translates a plurality of open reading frames in said DNA sequence into amino acid translations, and locates in each amino acid translation occurrences of said patterns to determine whether each said plurality open reading frames includes a putative gene ~~determines an open reading frame (ORF) in said DNA sequence.~~
3. (Currently amended) The system according to claim 2, wherein said patterns comprise biologically significant patterns of amino acids in amino acid sequences ~~processor generates an amino acid translation for said ORF.~~
4. (Currently amended) The system according to claim 1 ~~3~~, wherein said processor identifies a match of a pattern from said pattern database in said amino acid translation.
5. (Original) The system according to claim 4, wherein said patterns are derived from a parent database comprising at least one amino acid sequence.

Serial No. 10/059,421
Docket No. YOR920010126US2

3

6. (Original) The system according to claim 4, wherein said patterns are derived from a parent database comprising at least one amino acid sequence fragment.
7. (Original) The system according to claim 4, wherein said patterns are derived by using a pattern discovery algorithm.
8. (Original) The system according to claim 4, wherein said patterns are derived by using the Teiresias algorithm.
9. (Original) The system according to claim 4, wherein said ORF comprises a portion of said DNA sequence between a start codon and a stop codon.
10. (Currently amended) The system according to claim 4, wherein said processor reports said ORF ~~is reported~~ as a putative gene when a predetermined number of pattern matches is identified in said amino acid translation.
11. (Original) The system according to claim 4, wherein each pattern is assigned a weight depending upon a relevance of said pattern in determining whether said ORF comprises a putative gene.
12. (Original) The system according to claim 4, wherein said ORF is reported as a putative gene when the sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold.
13. (Original) The system according to claim 4, wherein said match is identified using a predetermined pattern matching algorithm.

Serial No. 10/059,421
Docket No. YOR920010126US2

4

14. (Original) The system according to claim 4, further comprising:
a memory device for storing data and instructions to be executed by said processor.
15. (Original) The system according to claim 4, further comprising:
a display device for displaying an output from said processor.
16. (Currently amended) A method of identifying genes, comprising:
providing a pattern database comprising patterns of amino acids;
determining an open reading frame (ORF) in a DNA sequence;
generating an amino acid translation for said ORF; and
determining whether said ORF includes identifying a putative gene in said DNA sequence
by locating ~~identifying~~ a match of a pattern from said pattern database in said amino acid
translation.
17. (Original) The method according to claim 16, wherein said pattern database is generated
from a database comprising at least one amino acid sequence.
18. (Original) The method according to claim 16, wherein said pattern database is generated
from a database comprising at least one amino acid sequence fragment.
19. (Original) The method according to claim 16, further comprising:
identifying said ORF as a putative gene when a predetermined number of pattern matches
is identified in said amino acid translation.
20. (Currently amended) The method according to claim 16, further comprising:
assigning a weight to each pattern depending upon a relevance of said pattern in
determining whether said ORF includes ~~is~~ a putative gene.

Serial No. 10/059,421
Docket No. YOR920010126US2

5

21. (Original) The method according to claim 16, further comprising:
displaying said match of said pattern in said amino acid translation.
22. (Original) The method according to claim 16, wherein said pattern database is generated using the Teiresias algorithm to derive said patterns from a parent database.
23. (Currently amended) A programmable storage medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method for identifying genes, said method comprising:
providing a pattern database comprising patterns of amino acids;
determining an open reading frames (ORF) in a given DNA sequence;
generating an amino acid translation for each ORF; and
determining whether said ORF includes identifying a putative gene in said DNA sequence
by locating identifying a match of a pattern from said pattern database in said amino acid translation.
24. (Previously presented) The system according to claim 3, wherein said processor determines for each pattern in said pattern database whether the pattern is present in said amino acid translation by locating instances of said patterns in said amino acid translation, until a sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold, at which point said processor identifies said ORF as a putative gene.
25. (Previously presented) The system according to claim 1, further comprising:
a parent database comprising a plurality of amino acid sequences, said patterns in said pattern database being derived from said plurality of amino acid sequences by using a pattern discovery algorithm;

Serial No. 10/059,421
Docket No. YOR920010126US2

6

a memory device for storing data and instructions to be executed by said processor; and
a display device for displaying an output from said processor.

26. (Currently amended) The system according to claim 25, wherein said ~~processor~~
~~determines an open reading frame (ORF) which~~ comprises a portion of said DNA sequence
between a start codon and a stop codon, ~~generates an amino acid translation for said ORF, and~~
wherein said processor identifies a match of a pattern from said pattern database in said
amino acid translation by using a predetermined pattern matching algorithm,
wherein each pattern is assigned a weight depending upon a relevance of said pattern in
determining whether said ORF comprises a putative gene, and
wherein said ORF is reported as a putative gene when either a predetermined number of
pattern matches is identified in said amino acid translation, or a sum of weights corresponding to
all patterns with matches in said amino acid translation exceeds a predetermined threshold.
27. (Previously presented) The system according to claim 1, wherein said processor accesses
said pattern database to retrieve said patterns from said pattern database.
28. (Previously presented) The system according to claim 1, wherein said processor is
electrically coupled to said input device and said pattern database.
29. (Currently amended) A system for identifying genes, comprising:
an input device which inputs a DNA sequence; and
a processor which:
accesses a pattern database comprising a plurality of patterns of amino acids;
translates an open reading frame (ORF) of said DNA sequence into an amino acid
translation; and
locates in said amino acid translation occurrences of said patterns from said

Serial No. 10/059,421
Docket No. YOR920010126US2

7

~~pattern database to determine whether said open reading frame includes , and locates instances of said patterns of amino acids in said DNA sequence to identify~~ a putative gene in said DNA sequence.

30. (New) A system for identifying genes, comprising:

an input device which inputs a query DNA sequence;

a processor which:

accesses a pattern database comprising a plurality of patterns of amino acids;

translates an open reading frame (ORF) of said DNA sequence into an amino acid translation; and

locates in said amino acid translation occurrences of said patterns from said pattern database to determine whether said open reading frame includes a putative gene in said DNA sequence; and

a display device for displaying an output of said processor, said output including an occurrence of said patterns in said amino acid translation,

wherein said patterns comprises patterns derived using a Teiresias algorithm,

wherein said open reading frame (ORF) comprises a portion of said DNA sequence between a start codon and a stop codon,

wherein said processor identifies a match of a pattern from said pattern database in said amino acid translation by using a predetermined pattern matching algorithm,

wherein each pattern is assigned a weight depending upon a relevance of said pattern in determining whether said ORF comprises a putative gene,

wherein said ORF is reported as a putative gene when either a predetermined number of pattern matches is identified in said amino acid translation, or a sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold, and

wherein after determining whether said open reading frame includes a putative gene, said processor analyzes a next open reading frame in said DNA sequence to determine whether said

Serial No. 10/059,421
Docket No. YOR920010126US2

8

next open reading frame includes a putative gene.